International Journal of Pediatric Otorhinolaryngology 124 (2019) 22–29

Contents lists available at ScienceDirect

International Journal of Pediatric Otorhinolaryngology

journal homepage: www.elsevier.com/locate/ijporl

Development of the Persian syntax comprehension test

Reyhane Mohamadi[a,*], Akram Ahmadi[b], Mehdi Dastjerdi Kazemi[c], Asghar Minaei[d], Zeinab damarchi[a]

[a] Department of Speech and Language Pathology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
[b] Department of Speech Therapy, School of Rehabilitation, Babol University of Medical Sciences, Babol, Islamic Republic of Iran
[c] Department of Psychology and Education of Exceptional Children, Allameh Tabataba'i University, Tehran, Iran
[d] Department of Assessment and Measurement, School of Psychology and Education, Allameh Tabataba'i University, Tehran, Iran

ARTICLE INFO

Keywords:
Comprehension
Persian
Reliability
Validity
Syntax

ABSTRACT

Objectives: Lack of formal tests to assess the various dimension of language in Persian is one of the main challenges of speech and language pathologists in Iran. The purpose of this study was to develop a Persian Syntax Comprehension Test to assess the syntax comprehension in Persian speaking children aged 4–10 years old.
Methods: the study included four phases. In the first phase, syntactic structures of Persian were extracted and then, items generation was performed. In the second phase, content validity was determined and images were designed for the items. In the third phase, two pilot studies were carried out and difficulty and discrimination indices for items were determined and in the last phase, 788 typically-developing children (436 children aged 4–6 and 352 children aged 6–10 years old) and 15 children with Developmental language disorder were recruited then, psychometric properties (construct and concurrent validity, test-retest, and split-half reliability) were evaluated. In the final step, standard score and percentile were calculated. We used Statistical Package for the Social Sciences, version 24.0 (SPSS, Inc., Chicago, IL) for the statistical analysis of the data. The significance level was ($P < 0.05$).
Results: Final version of Persian syntax comprehension test contains 24 syntactic structures and 96 items. Items' Difficulty indices were between 0.33 and 0.90. There was a significant difference among the age groups for the mean total score of the Persian syntax comprehension test. In addition, a high correlation existed between total scores of this test and those of grammar understanding subtest of Test of langue development ($r = 0.52$, $P < 0.001$). The correlation of total Persian syntax comprehension test score of the two rounds of test performance and the split-half coefficient were estimated to be 0.56 and 0.85, respectively.
Conclusion: It seems that the Persian Syntax comprehension test has satisfactory values for the reliability and validity measures, and it can be used as a suitable instrument by researchers and clinicians.

1. Introduction

Sentence comprehension especially about complex ones plays a significant role in children's academic achievement and social development [1]. Language disorders, particularly comprehension disorders, have lasting effects including adverse personal, familial, and social conequences [2,3]. In addition, comprehension deficits put children about to begin formal education at risks of inability to develop reading and writing [4,5]. Therefore, assessing langauge comprehension, though being a routine task performed by speech and langauge pathologists on a daily basis, is a demanding and challenging issue. On the other hand, determining basic levels of language performance in children with langauge disorders through formal and informal assessments is an indispensable task for speech and language pathologists. This can reveal children's strengthes and weaknesses, and if necessary, help set a baseline for linguistic interventions.

In clinical setting, diagnosis and clinical decisions are often made based on norm-referenced tests. The lack of specialized tests in Persian has made Persian speech-language pathologists utilize informal assessments for conducting most of their clinical activities and research. complete reliance on informal assessments, generally depends on degree of practitioner's clinical skill and experience [6]. This can lead to improper clinical judgments and inaccurate research findings. Obviously, the use of norm-referenced tests reduces the effect of personal judgments on the preformation, and scoring and interpretation of scores. Therefore, clinical conclusions and decisions are also made

* Corresponding author.
E-mail address: mohamadi.r@iums.ac.ir (R. Mohamadi).

https://doi.org/10.1016/j.ijporl.2019.05.032
Received 24 November 2018; Received in revised form 19 April 2019; Accepted 23 May 2019
Available online 27 May 2019

solely on the basis of individual judgment and experience. In addition, an accurately-conducting research work requires formal and norm-referenced tests. To conduct research and/or make clinical decisions, which includes diagnosis, evaluation, determining the nature of the disorder, designing an intervention plan, and measurement of the progress in intervention, access to valid and reliable tools is essential.

The developmental assessment tools of language and syntactic comprehension which are available in English contain sentences and issues that are not compatible with the Persian culture and linguistic structure and thus, the direct translation or back translation and adaption of international language tests into Persian will not bridge the existing gap. This necessitates creating a tool which is by no means suitable for assessing native Persian children with Iranian culture. The sentence structures have special features in Persian and these characteristics need to be assessed by language-specific measures. Therefore we cannot use English test for the assessment of syntax comprehension in Persian speaking children.

Undoubtedly, test adaptation from one language to another is scientific and even have benefits such as facilitating intercultural and interdisciplinary comparative studies, saving time and costs [7,8]. However, nobody can simply assume that the tools which are developed to measure the concepts and structures of one language and culture, can easily and with a little change in the vocabulary, be used in other languages and cultures [9]. The comparability of scores and measures resulting from the adaptation of instruments and tests depends on their validity and their level of equivalence [8,9]. One of the requirements for the use of tests belonging to another language and culture is to equate its concepts and structures. However, some interdisciplinary researchers have acknowledged that a high percentage of research in this area lacks due to the weakness in testing compliance and inequality of structures [10]. Therefore, there is always a concern that the version of a valid test in another language which is practically equally assumed, does not fit into each other, and only a general pattern of the original test remains. This danger increases more when the language itself is the subject of study and testing. Theoretical studies have also shown that culture has an impact on children's development, particularly on their social development [11], but certainly taking advantage of the existing international tests general framework could be useful. Unique features of the Persian language demand that a completely native test be built in accordance with its features.

This study was part of a larger project. The first part which was published consisted of generation and content validation of the Persian Syntax Comprehension Test (PSCT). The aim of the present study was to determine its psychometric features based on the characteristics of the Persian language for children aged 4–10 years old. All the analyses in this study were carried out in two age groups, 4–6 and 6–10 years.

2. Material and method

A methodological, cross sectional and descriptive study was carried out to develop the Persian syntax comprehension test.

2.1. Participants

Two groups of children participated in this study; the first group included 788 typically developing children (TD children) (436 children aged 4–6 and 352 children aged 6–10 years). The second group included 15 children with Developmental language disorder (DLD).

To select the TD children aged 4–6, the city of Tehran was first divided into three geographic regions of the north (including 1–4 areas), the center (including areas 14–5, 21 and 22) and the south (including 15–19). Then 3 areas from the northern geographic region (areas 2, 3, 4), seven regions from the geographical region of the center (areas 5, 6, 7, 8, 10, 11, 13) and two areas from the southern geographic region (areas 15 and 18) were selected in a simple random manner. Thus, 12 areas of Tehran were selected in this part of the study.

Table 1
The sample sizes of the groups across geographic areas, age, and gender.

N	Age(M,SD)	gender	north	center	south	total
Group1	4; 0–4; 5(4; 3,0.15)	boys	10	18	14	42
		girls	13	29	15	57
		subtotal	23	47	29	99
Group2	4; 6–4; 11(4; 7,.13)	boys	21	38	7	66
		girls	15	28	12	55
		subtotal	36	66	19	121
Group3	5; 0–5; 5(5; 2,.15)	boys	10	38	4	52
		girls	11	19	20	50
		subtotal	21	57	24	102
Group4	5; 6–5; 11(5; 7,.14)	boys	10	41	20	71
		girls	2	29	12	43
		subtotal	12	70	32	114
Group5	6; 0–6; 11(4; 3,0.15)	boys	5	14	5	24
		girls	4	4	11	19
		subtotal	9	18	16	43
Group6	7; 0–7; 11(4; 3,0.15)	boys	5	27	13	45
		girls	13	18	21	52
		subtotal	18	45	34	97
Group7	8; 0–8; 11(4; 3,0.15)	boys	5	43	15	63
		girls	3	12	21	36
		subtotal	8	49	33	99
Group8	9; 0–9; 11(4; 3,0.15)	boys	12	49	11	72
		girls	9	14	18	41
		subtotal	21	63	29	113

M:mean, SD: standard deviation.

Of these 12 areas, 42 educational centers (31 kindergartens and 11 preschool centers) were selected and 436 children were selected randomly from these centers. In order to select children from 6 to 10 years of age, after dividing Tehran into three geographical regions north, south and center, from the north geographical region, areas of 1,2,3, from the center region, areas 5,6,7, 10,12 and from the southern region area of 15 were randomly selected. Given that the number of areas in the center region was greater than the north and south (12 in the central area versus 4 areas in north and 4 in the south), the number of selected areas was higher and the sample size selected from the center region was 2.5–3.5 times more than the other two regions. Table 1 shows the sample sizes of the groups according to geographic areas, age, and gender.

Fifteen children with DLD who aged 4–7 years old participated in this study. All children with DLD were included based on the following criteria; they passed a hearing screening and an oral motor screening examination. Each child with DLD scored above 85 on the Persian version of Wechsler preschool and primary scale of intelligence (WPPSI) [12]. Children with DLD had scores of at least a 1.5 standard deviation below the mean in Persian version of test of language development –Primary3 (TOLD-P:3) [13].

For all suspected children with DLD, the TOLD-P: 3 test was fully performed to calculate their linguistic score and ensure that they scored at least 1.5 deviations below the average. Then considering other criteria mentioned above, diagnosis of DLD was made. However, in order to evaluate the construct validity in the study, considering that only the sub-test of the grammatical understanding was similar to PSCT, the correlation of this sub-test was calculated with the total score of the PSCT.

2.2. Instruments

2.2.1. TOLD-P3

TOLD-P3 is standardized on children with age range of 4 years old to 8 years and 11 months in Tehran. Six sub-tests are used for evaluating the semantics and syntax and include core sub-tests: picture vocabulary, relational vocabulary, oral vocabulary, grammatical understanding, sentence imitation, and grammatical completion. The first three sub-tests evaluate semantics, and the second 3 sub-tests evaluate

syntax. The only sub-test that is allocated to syntax comprehension is the test of grammatical understanding, which measures children's ability to comprehend sentences with 25 items. In the Persian version of the test, Cronbach's alpha coefficients for the sub-tests ranged between 0.74 and 0.94. In this study, six core sub-tests were performed on children with DLD.

2.2.2. Persian version of Wechsler pre-school and primary scale of intelligence

The Persian version of Wechsler pre-school and primary scale of intelligence (WPPSI) was used for measuring the intelligence quotient of children 4–6 years and 6 months and includes verbal and non-verbal parts. Reliability values of the test were reported between 0.27 and 0.90. In this study, the non-verbal part of the test, including the sub-tests of animals' house, picture completion, Mazes, geometric design; cubes were used to measure the examinees' intelligence. The test was performed on DLD children.

2.2.3. Persian Syntax Comprehension Test (PSCT)

Development of PSCT is the main objective of this study. This test consists of 24 syntactic structures and there are 4 items of each structure. With 96 items in total, it measures the syntax comprehension of Persian-speaking children aged 4–6 years old. The total PSCT score equals the total number of syntactic structures to all 4 items of which the correct answer is given. The scores of individuals in the test can range from 0 to 24.

2.2.4. Test of syntax comprehension image booklet

The images designed for the test were prepared in the form of a booklet to assess syntactic comprehension in samples of the study.

3. Test development

3.1. Phase 1: generation of syntax structures and items

To extract syntactic structures, deductive-inductive approach was used [14]. Finally, 198 items for 33 syntactic structures using 108 content words were prepared. The first phase of this work has been published previously in which the details of generation of these items has been described [14].

3.1.1. Phase 2: content validity

Content validity is the extent to which a test is connected and related to what it is going to assess. In order to determine the content validity of the structures and items, content validity ratio (CVR) originally proposed by Ayre and Scally was used [16]. In the current study, 14 experts reviewed structures and 12 experts reviewed the items. Selection of experts was based on their specialized fields, the history of their research, and their clinical experience. Item-by-item analysis of experts' ideas was conducted by a research team. Decisions about the removal, modification or replacement of new items were made. The minimal values of CVR for syntactic structures were 0.47. Therefor 24 structures were selected. 120 items were selected on the basis of their CVR (CVR > 0.50).

3.1.2. Phase 3: drawing images

120 colorful four-item pictures were drawn based on Persian culture, two images on the top of the page and two on the bottom of it. The images were drawn such that the characters in them were interesting for children. Selection of the target and distractor images was of particular importance. Distractor pictures were semantic or syntactic or both, depending on the item. Images of 34% of the items contained at least one semantic distractor image. (items belong to syntactic Structures below had one to three semantic distractors reversible SOV, intransitive basic sentence, simple negative sentences, transitive active simple sentences, comparative adjectives, passive sentences, free pronouns, bound pronoun of verbs, superlative adjectives were semantic) in the rest of the items, there were three syntactic distractors. Designing images in this way, although making the test more difficult and complicated, also reduced the chance of accidental responses.

All images were prepared by a graphic artist. The necessary modifications including increasing clarity, changing the position, color of the target image and three confounding images were made. After three stages of modifications, 5 speech and language pathologists commented on the resolution of the images in three words, it is clear, relatively clear and unclear. Finally, images agreement coefficient was calculated using Statistical Package for the Social Sciences, version 16.0 (SPSS, Inc., Chicago, IL) (SPSS16) (Contingency coefficient 0.77 $p < 0.001$).

3.1.3. Phase 4: pilot studies

3.1.3.1. Pilot study 1

As previously mentioned, the content validity of 120 items was approved and the images were prepared for them. The first pilot study was carried out to analyze the quality of items and images, where 30 children aged 4–6 years participated.

The purpose of this pilot study was to examine the quality of the items and images so that children were not given the correct or incorrect answers. In this section, in addition to examining the children's responses to the items, their verbal statements were also recorded during the response to the items. The main goal of this step was to identify difficult vocabulary for children and replace by appropriate items. Another goal was refining items if necessary and replacing vague and inappropriate images with appropriate images.

After administration of the final version of the test on the participants, qualitative analysis of the correct and incorrect answers was performed by authors. The comprehension of two words (one verb and one noun), affected sentence comprehension. These two words were replaced by two appropriate ones. The images related to 30 items were modified, with modifications including removing distracting details, color and color intensity, the position of hands, legs, and direction of characters' looks to better display the action expressed by verbs, and match the size of the four images related to an item. Thus, six items were removed which were later replaced by clearer images to which children gave more correct answers. Ultimately, 114 items were left for use in the test.

3.1.3.2. Pilot study 2

In the second pilot study, 100 children aged 4–6 years old were tested. The multi-stage stratified method was used. At first, Tehran was divided into three geographic districts, north (including districts 1–4), center (including districts 5–14, 21 and 22), and south (including districts 15–19). In the next step, 8 districts were randomly selected. Among the kindergartens and preschool centers of the districts, 7 kindergartens and 4 pre-school centers were selected using simple randomization. All 100 children's first language was Persian and had a history of normal hearing thresholds. They had no history of delay in psychomotor milestones. 46% male and 54% female, 20% were in the age range of 4 to 4; 5, 18% were in the age range 4; 6–4; 11, 30% aged 5; 0–5; 5 and 32% in the age range of 5; 6–5; 11, respectively. After the second pilot study, the difficulty index was calculated by the percentage of respondents who answered the items correctly [17]. All syntax structures were arranged in an ascending order of difficulty. Discrimination index was rated using the point biserial correlation coefficient between each item score and the total PSCT score [17]. After the second pilot study, the total number of items was reduced from 114 to 96.

3.1.4. Phase 5: validity and reliability analysis

3.2. Construct validity

In addition to the content validity which results were stated in Section of construct validity was also assessed. Construct validity was examined based on the two evidences: heterogeneous group comparisons, and age group developmental trend. Several studies demonstrated that DLD children perform poorly in the comprehension of syntactic structures compared to children of the same age and failure to comprehend the grammar is a prominent feature of children with language disorders [18–21]. The performance of TD and DLD children in each syntactic structures was compared using a one-way multivariate analysis of variance (MANOVA) and alpha Bonferroni correction was applied to correct for multiple testing. To compare TD and DLD children in PSCT total score, independent-sample T test was used Univariate analysis of variance (ANOVA) was used to examine total PSCT and MANOVA was used for each syntactic structure score across the age groups to test the differences among the age groups.

3.3. Concurrent validity

Concurrent validity was determined based on Pearson's correlation coefficient between the total PSCT score and subtest of grammar understanding of TOLD was calculated.

3.4. Reliability

Internal consistency was evaluated by totaling odd and even syntactic structures and calculating the correlation between them. To examine Test-retest reliability, 48 subjects with an average age of 66.06 months were randomly selected. Mean interval between the two rounds was 18 ± 4 days. Pearson's correlation coefficient was rated comparing rounds 1 and 2 of total PSCT score.

3.4.1. Phase6. Conversion of raw scores into standard scores

The Total raw score of the test for each child was between 0 and 24. If a child obtained a score of 0 it means that he/she did not pass every 4 items of any of the 24 syntactic structures. Score 24 means that the answers were correct for every 4 items of all syntactic structures. In this study, raw scores were transformed to standard scores and percentiles (Table 6). Standard scores have an approximately normal distribution which makes them suitable to use in statistical analysis.

4. Results

4.1. Difficulty and discrimination analysis

Difficulty and discrimination indices at the final stage in which the test was conducted on 788 children, were recalculated. Table 2 shows the difficulty index average of the items related to each structure and the items associated with the mode of discrimination index of each syntactic structure based on the normative sample. Difficulty indices were between 0.33 and 0.90.

4.2. Validity analysis

Construct validity was measured through heterogeneous group differences and developmental differences. MANOVA was used to estimate heterogeneous group differences. DLD and TD children were considered as independent variables and the syntactic structures were considered as dependent variables. The results showed that TD children in all syntactic structures, except for the 3 structures (Table 3), had better performance than the DLD children, and the differences were statistically significant. Hoteling's Trace = 0.65 F [24,426] = 4.211 P < 0.001 (partial eta squared = 0.19). A significant difference was observed in total PSCT score between TD and DLD children (P < 0.001). To investigate the developmental trend, each syntactic structure score across age groups, using MANOVA, was compared (Tables 4 and 5). It was found that comprehension of syntactic structures with increasing age was gradually increased in 4–6 year old children (Pillai's Trace = 0.37 F [24, 1233] = 2.41, p < 0.05 partial eta squared = 0.12). Post hoc comparisons with Bonferroni correction, P-values smaller than 0.002 (0.05/24), showed that 4; 0–4; 5 year old children and 5; 5-5; 11 year old children were significantly different in all syntactic structures (Table 4). In 6–10 year old children Pillai's Trace = 0.98 F [24, 837.22] = 2.41, p < 0.051 partial eta squared = 0.12. Post hoc comparisons showed that 6–10 year old children were significantly different in comprehending ten structures. 6; 0–6; 11 years old children and 9; 0–9; 11 years old children were significantly different in comprehending seven structures 7; 0–7; 11 years old children and 9; 0–9; 11 years old were significantly different in comprehending nine syntax structures (Table 5).

To compare the total PSCT score in 4–6 age groups one-way ANOVA was used. ANOVA test results showed a significant difference in the total PSCT score across age groups. F (3,432) = 34.62 P < 0.001. Post hoc age group comparisons, with a Bonferroni correction for multiple comparisons were applied. The performance of 4; 0–4; 6 years old TD children was poorer than the performance of 4; 6–4; 11, 5; 0–5; 5 and 5; 5-5; 11year o0ld TD children. The total PSCT score of 4; 6–4; 11 year old and 5; 0–5; 5year old children were not significantly different, although the total score of 5; 0–5; 5 year old group was more than the total score of 4; 6–4; 11 year old group. concurrent validity demonstrated showed that there was a moderate correlation between total scores of PSCT and the scores of grammar understanding subtest of Told (r = 0.52).

4.3. Reliability analysis

Person's correlation coefficient was used to measure Test-retest Reliability. All correlations were significant at 0.01 except for a syntax structures. Pearson Correlation coefficients ranged from 0.33 to 0.85, respectively. The correlation of total PSCT score of the two rounds was estimated 0.56. Split-half reliability was calculated as a measure of internal consistency. The split-half coefficient was estimated 0.85.

4.4. Conversion of raw scores into standard scores analysis

Raw score transformed to Z scores and then Z scores were converted to standard scores with an average of 100 and a standard deviation of 15. If a child is in the age range of 4; 0–4; 6 and passes 2 to 6 syntactic structures, he will fall into the normal range. A child aged between 4; 6–4; 11 must pass at least 4 syntactic structures in order to be in normal range according to his age. Similarly, the minimum raw score for the age group of 5; 0–5; 5 and 5; 5-5; 11, are 4 and 7 respectively. For children in the age range of 6; 0 to 9; 11 the normal range for the raw score is between 12 and 18 (see Table 6).

4.5. Other finding

Total scores of participants in PSCT were calculated and compared in two age groups in terms of gender. As shown in Table 7 both two groups did not show significant difference for the total score of PSCT in terms of gender.

5. Discussion

The aim of this study was to develop a reliable and valid test of syntax comprehension that specifically address the difficulty encountered in interpreting syntactic constructs. A total of 33 syntax and 198 items were generated. During the content validity process, 24 syntactic structures and 120 items remained. The first pilot study led to a decrease in the items from 120 to 114.after calculating the difficulty and discrimination index, 96 items for 24 syntactic structures remained.

The current study demonstrated strong construct validity. The results showed that with the increase in age, the mean score of each

Table 2
Difficulty and discrimination analysis for the items of the Persian syntax comprehension test.

Syntax structures	4–6 year old		6–10 year old	
	Difficulty	discrimination	Difficulty	discrimination
intransitive basic sentence	0.93	0.12	0.97	0.41
simple negative sentences	0.89	0.34	0.87	0.44
transitive active simple sentences	0.83	0.28	0.85	0.50
prepositional phrases	0.73	0.30	0.86	0.49
reversible SOV	0.75	0.35	0.86	0.48
positive conjunction in compound sentences	0.78	0.31	0.78	0.49
sentences with noun coordinated phrase	0.76	0.35	0.80	0.44
superlative adjectives	0.71	0.29	0.72	0.50
A phrase not B phrase	0.69	0.23	0.70	0.46
locative adverbs	0.67	0.29	0.77	0.46
comparative adjectives	0.66	0.39	0.61	0.44
bound pronoun of verbs	0.65	0.36	0.62	0.17
passive sentences	0.63	0.36	0.63	0.40
negative conjunction in compound sentences	0.58	0.29	0.45	0.35
omitted subject in compound sentences	0.55	0.46	0.58	0.36
Pronoun binding	0.54	0.36	0.56	0.43
adjective genitive sequences	0.54	0.20	0.42	0.31
subject relative clauses	0.35	0.30	0.43	0.43
direct object relative-clauses	0.49	0.28	0.53	0.35
tense-aspect-mood of verbs	0.46	0.24	0.28	0.24
omitted object in compound sentences	0.47	0.22	0.49	0.25
free pronouns	0.44	0.26	0.41	0.31
subject-verb agreement	0.38	0.20	0.28	0.17
two object verbs	0.33	0.23	0.33	0.26

Table 3
Mean and standard deviation of syntactic structures and total SCT scores for the typically developing children and children with DLD (evidence of the construct validity).

Structures	DLD(N = 15)		TD(N = 436)		P
	M	SD	M	SD	
1	3.07	0.70	3.75	0.53	.000
2	2.40	1.50	3.58	0.79	.000
3	2.40	1.10	3.34	0.81	.000
4	1.87	1.10	2.94	1	.000
5	2.00	1.10	3.01	1	.000
6	2.53	0.91	3.15	0.92	.011
7	1.13	0.99	3.07	1.1	.000
8	1.87	1	2.85	1	.000
9	1.27	1.10	2.75	1	.000
10	1.93	0.96	2.71	1	.006
11	0.87	0.74	2.67	1.2	.000
12	1.13	1.1	2.62	1.1	.000
13	1.20	0.94	2.54	1.1	.000
14	1.53	1.1	2.35	1.1	.006
15	0.67	1	2.25	1.4	.000
16	0.80	0.67	2.19	1.1	.000
17	0.67	0.9	2.17	1	.000
18	1.13	0.91	2.14	1.1	.000
19	1.40	1.1	1.97	1.1	.063
20	0.93	1	1.85	1	.000
21	1.07	0.70	1.91	1	.000
22	1.00	0.65	1.76	1.1	.010
23	1.20	0.86	1.53	0.98	0.202
24	1.00	1	1.32	1	0.232
Total score	1.40	1.80	6.63	4.30	.000

syntactic structure increased. Children of the 9-9; 11 scored higher in all syntax structures than the other age groups. This finding, in addition to being an evidence of the construct validity, confirms the development of syntax comprehension from 4 to 10 years of age. Considering the mean of the total score of PSCT or 4–6 years old children, the result indicated that all age groups except the second age group (4; 6–4; 11) and the third (5; 5-5; 11) had a significant difference in total score of the PSCT. A possible explanation for this might be that they are close to their average age range. The average distance between these two groups was about 4 months, while the average age range was between the first age group (4; 0–4; 5) and the second (4; 6–4; 11), the third (5; 0–5; 5) and the fourth (5; 6–5; 11) were about 6 months. Considering the mean of the total score of PSCT for 6–10 years old children, there was a significant difference between the age groups in the total score of the test except for the age group 6; 0–6; 11 vs. 7; 0–7; 11, and 7; 0–7; 11vs 8; 0–8; 11. Construct validity assessment using heterogeneous group differences showed TD children compared to DLD children had a much higher PSCT score, and this difference was statistically significant. The results showed that regarding all syntactic structures (apart from three structures) TD children outperformed DLD children. TD children also obtained low scores in comprehending those three structures. Grammatical limitations are one of the key symptoms of developmental language disorders [21]. According to Vander Lilly et al. [22], there is at least one group of children with specific language impairment that show a more severe difficulty in the syntactic understanding than the semantics and phonological domains. The problems of these children show themselves further when semantic and pragmatic cues are eliminated [23].

Considering the concurrent validity the correlation coefficient between the score of the sub-test of grammar comprehension and the total score of the syntactic comprehension test was moderate. This correlation value shows these two scales measures almost one behavior. This result may be explained by the fact that the PSCT has structures and items that do not exist in the sub-tests of grammatical understanding of TOLD-P3. The sub-tests of grammatical understanding assess only 25 sentences. In addition, one of the criticisms of the subtests is related to the selection of sentences and images. Although the purpose of this subtest is to understand grammar, some of the items of this subtest precisely do not measure the syntax comprehension and knowing the meaning of one or two words in these items is sufficient to select the correct image. Therefore, the wrong answer to these terms does not mean the lack of understanding of grammatical structures [24]. In PSCT, we tried to eliminate this limitation as far as possible by generating reversible structures. Considering the above reasons, the correlation coefficient obtained between the PSCT and the sub-test of grammatical understanding of TOLD-P3 can be justified. The current study demonstrated high values for test-retest. There was a high

Table 4
Mean and standard deviation of syntactic structures and total CT scores for 6–10 aged groups.

structures	4; 0–4; 5 (N = 99)		4; 6–4; 11 (N = 121)		5; 0–5; 05 (N = 102)		5; 06–5; 11 (N = 114)		Post hoc comparisons[a]
	M	SD	M	SD	M	SD	M	SD	
1	3.53	0.70	3.71	0.52	3.78	0.45	3.87	0.34	b,c
2	3.16	1.00	3.44	0.82	3.74	0.45	3.83	0.31	b,c,f
3	2.90	0.91	3.13	0.89	3.52	0.52	3.69	0.50	b.c.d,f
4	2.58	1.00	2.81	1.00	3.03	0.65	3.22	0.90	b,c,f
5	2.60	1.10	2.93	1.00	3.09	0.94	3.30	0.93	b,c,d
6	2.81	0.96	3.04	0.91	3.09	1.00	3.55	0.71	c,e,f
7	2.62	1.20	2.91	1.10	3.06	0.99	3.41	0.79	b,c
8	2.39	1.00	2.65	1.00	2.93	1.00	3.27	0.74	b,c,e,f
9	2.38	1.00	2.67	1.00	2.85	0.94	2.89	0.93	a,b,f
10	2.25	1.00	2.65	1.00	2.80	1.00	3.01	0.99	b.c.d,f
11	2.12	1.20	2.33	1.30	2.85	0.95	3.15	0.96	a,b,c
12	2.08	1.30	2.37	1.20	2.84	1.10	3.01	0.88	b,c,e,f
13	2.06	1.20	2.26	1.10	2.68	1.00	3.02	1.00	b,c,e,f
14	2.10	1.10	2.14	1.00	2.35	1.10	2.70	1.00	b,c,f
15	1.52	1.40	1.86	1.30	2.50	1.00	2.86	1.10	c,f
16	1.74	1.10	2.00	1.00	2.25	1.30	2.58	1.10	b,c,f
17	1.75	1.00	2.02	1.00	2.28	1.1	2.43	0.90	b.c.d,f
18	1.66	1.10	1.95	1.00	2.11	0.99	2.69	1.10	b,c,f
19	1.77	1.00	1.74	1.00	2.00	1.10	2.31	1.20	c,f
20	1.66	1.00	1.66	1.00	1.86	1.10	2.09	1.00	c,f
21	1.67	1.00	1.83	1.00	1.89	1.00	2.09	0.99	c
22	1.37	1.00	1.72	1.10	1.81	1.00	2.03	1.10	b,c
23	1.29	0.94	1.46	0.93	1.59	0.99	1.73	0.99	c
24	1.04	0.94	1.28	0.90	1.35	1.10	1.56	1.10	c
Total score	4.16	3.29	5.78	3.51	6.92	4.197	9.41	4.47	a,b,c,e,f*

P-value: * < 0.001.
Comparisons of 4–6 aged groups: a:4; 0–4; 5 vs. 4; 6–4; 11, b: 4; 0–4; 5 vs.5; 0–5; 5, c: 4; 0–4; 5 vs. 5; 5-5; 11 d: 4; 6–4; 11 vs. 5; 0–5; 5, e: 5; 0–5; 5 vs. 5; 5-5; 11 f: 4; 6–4; 11 vs. 5; 5-5; 11.

Table 5
Mean and standard deviation of syntactic structures and total CT scores for 6–10 aged groups.

structures	Age groups								
	6 -6; 11(N = 43)		7-7; 11(N = 97)		8-8; 11(N = 99)		9-9; 11(N = 113)		Post hoc comparisons[a]
	M	SD	M	SD	M	SD	M	SD	
1	3.95	0.21	3.95	0.52	3.97	0.17	3.99	0.09	
2	3.86	0.35	3.90	0.33	3.81	0.39	3.90	0.29	
3	3.63	0.61	3.79	0.30	3.84	0.42	3.94	0.24	
4	3.81	0.39	3.79	0.45	3.82	0.38	3.95	0.22	
5	3.70	0.59	3.75	0.45	3.87	0.36	3.93	0.25	
6	3.42	0.85	3.63	0.56	3.71	0.55	3.85	0.50	
7	3.58	0.69	3.73	0.69	3.73	058	3.83	0.44	
8	3.63	0.53	3.56	0.53	3.69	0.52	3.80	0.74	
9	3.49	0.76	3.41	0.64	3.64	0.59	3.81	0.47	e
10	3.47	0.76	3.70	1.80	3.51	0.87	3.81	0.49	
11	3.12	1.00	3.30	0.64	3.46	0.81	3.62	0.69	
12	2.95	0.87	3.38	0.86	3.54	0.77	3.72	0.55	a,b.c,e
13	3.12	0.93	3.36	0.82	3.35	0.90	3.76	0.53	c,e
14	2.67	1.10	3.07	0.81	3.07	1.00	3.34	0.91	
15	3.44	0.73	3.34	0.94	3.41	0.78	3.55	0.66	
16	3.19	0.88	3.14	0.88	3.31	0.96	3.59	0.63	
17	2.72	1.10	2.98	1.05	3.04	1.00	3.35	0.84	e
18	2.58	1.10	3.05	0.91	3.00	0.99	3.29	0.97	
19	2.67	1.10	2.99	1.00	3.32	1.00	3.55	0.74	b,c,e
20	2.37	1.00	2.73	1.00	2.52	1.00	3.12	0.87	c,f
21	2.63	1.20	2.91	1.00	3.10	1.00	3.50	0.82	c,e
22	2.28	1.10	2.80	1.10	3.14	0.86	3.26	0.99	c,e
23	2.12	1.20	2.34	1.10	2.49	1.20	3.10	0.93	d,e,f
24	2.02	1.10	2.30	1.30	2.49	1.20	3.19	1.00	c,e,f
Total score	11.58	5.19	13.39	5.03	14.5	4.81	17.47	4.30	b,c,e,f

P-value: < 0.05.
Comparisons of 4 age groups: a:6; 0–6; 11 vs. 7; 0–7; 11, b: 6; 0–6; 11 vs.8; 0–8; 11, c: 6; 0–6; 11 vs. 9; 0–9; 11 d7; 0–7; 11vs. 8; 0–8; 11, e: 7; 0–7; 11 vs. 9; 0–9; 11 f: 8; 0–8; 11 vs. 9; 0–9; 11.

Table 6
Conversion of structures into standard scores.

	Number of syntactic structured passed																								
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
4;0 – 4;5	81	86	**90**	**95**	**99**	**104**	**108**	**113**	**117**	122	127	131	136	140											
4;6 – 4;11	75	80	84	88	92	97	101	105	109	114	118	122	127	131	135										
5;0 – 5;05	75	79	82	86	**90**	**93**	**97**	**100**	**104**	**107**	**111**	**115**	**118**	122	125	129		136	140						
5;06 – 5;11	68		78		82	85	89	**92**	**95**	**99**	**102**	**105**	**109**	**112**	**116**	**119**	122	125	129	132					
6 – 6;11			**61**			**70**	**73**	**76**	**79**	81	85	88	91	94	96	99	102	105							
7 – 7;11				64	**67**		**73**	**76**	**79**	**82**	**85**	88	91	94	96	99	102	105	108	111	114	117	120		126
8 – 8;11					**67**		73	76	79	82	85		91	94	97	99	102	105	108	111	114	117	120		
9 – 9;11	-				67					82	85	88	91	94	97	99	102	105	108	111	1140	117	120	122	126

Table 7
Comparison of boys and girls in administration of PSCT in age groups.

4–6 year old				P-value	6–11year old				P-value
Girls		Boys		0.71	Girls		Boys		0.56
Mean	SD	Mean	SD		Mean	SD	Mean	SD	
7.55	4.63	6.55	4.27		14.97	5.34	14.66	5.03	

correlation between round 1 and 2 of test performance. This finding may imply that PSCT measures the syntax comprehension stably [15].

In everyday life, syntax is just one of the sources of information used to get the meaning of the utterances. In language comprehension, Bishop [20] described a wide range of underlying problems that can include poor listening discrimination, limited knowledge of vocabulary, weak verbal memory, grammatical difficulty, slow speech processing, difficulty in inferring the meaning of the context, Weak social cognition, and insensitivity to nonverbal cause. PSCT By trying to minimize the impact of factors that are effective in understanding puts the child in an unnatural situation. If one wants to identify the child's comprehension in everyday life, PSCT is not a good option. A person who has poor performance in PSCT may have a remarkable comprehension in everyday life by using word and nonverbal cues. In PSCT syntax processing demand is much more than what is normally needed in everyday life, because the test is developed to minimize redundancy, so each word and the order of its occurrence are important in the items.

PSCT can identify children with impairment in syntactic comprehension. But test developers do not claim at all that the test is a diagnostic tool for children with language impairment. This is because to identify children with language impairment and also to give an overall profile of strengths and weaknesses of examinees, a comprehensive test is needed to cover multiple domains of language, But this tool can contribute to the evaluating, treatment planning and making clinical decisions considering weaknesses and characteristics of children with language disorders-besides informal assessments in clinical settings and also puts an end to the mere use of informal assessments in Iran. In Persian language we had no specific information regarding the developmental process of comprehension in syntactic structures; it seems that one of the most important findings of this study was identifying more difficult syntactic structures in the Persian language. Girls and boys had not significant difference in total scores of PSCT. Hyde and Linn discussed in the meta-analysis that gender involves small variance (10–15%) in speech skills of children. We had some limitations in our study that should be mentioned [25]. One of the limitations of this study was that no larger sample size was provided which can offer the possibility of conducting factor analysis. Second, only 15 children with DLD were recruited for the discriminant validity. Including the larger sample size of children with DLD undoubtedly increase the clinical validity of this scale. Third, it was better to use different groups of children with language disorders (different types of hearing impaired children and children with language disorders due to cognitive deficits), as evidence of construct validity. Finally, future research should be performed to evaluate the psychometric properties of this instrument in adolescents and adults.

6. Conclusions

PSCT is a new test in the Persian language proved to have an adequate reliability and validity. Therefore, we expect the PSCT as a suitable tool meets some parts of research and clinical needs of speech-language pathologists.

Acknowledgements

This study on children aged 4–6 years old was supported by the University of Social Welfare and Rehabilitation Sciences science with ethical code (Ref:USWR.REC.1393.10). The study on children aged 6–10 years old was supported by Iran University of Medical Sciences with ethical code(Ref:94-05-32-27396). Authors of this article do appreciate all professors and experts that helped us by their scientific opinions especially Dr. Hossein Samei, Dr.Mohammad Reza Razavi and Dr.Shahin Nematzadeh. We would like to thank the children who participated in the study, and their parents.

References

[1] D.M. Gummersall, C.J. Strong, Assessment of complex sentence production in a narrative context, Lang. Speech Hear. Serv. Sch. 30 (1999) 152–164.
[2] J. Coplan, Evaluation of the child with delayed speech or language, Pediatr. Ann. 14 (1985) 203–208.
[3] P.A. Silva, S. Williams, R. McGee, A longitudinal study of children with developmental language delay at age three: later intelligence, reading and behaviour problems, Dev Med Child Neurol 29 (1987) 630–640.
[4] S. Ebbels, J. Dockrell, H.K.J. van der Lely, Production of Change of State, Change of Location and Alternating Verbs: a Comparison of Children with Specific Language Impairment and Typically Developing childrenLanguage and Cognitive Processes, (2012).
[5] C. Miniscalco, M. Westerlund, A. Lohmander, Language skills at age 6 years in Swedish children screened for language delay at 2(1/2) years of age, Acta Paediatr. 94 (2005) 1798–1806.
[6] K. Shipley, J. McAfee, Assessment in Speech-Language Pathology: A Resource Manual, Cengage Delmar Learning, City, 2009.
[7] R.K. Hambleton, Translating achievement tests for use in cross-national studies, Eur. J. Psychol. Assess. 9 (1993) 57–68.
[8] F. Van de Vijver, K. Leung, J.W. Berry, Y. Poortinga, J. Pandey (Eds.), Methods and Data Analysis of Comparative Research, second ed., Allyn & Bacon, City, 1997, pp. 257–300.
[9] C.H. Hui, H.C. Triandis, Measurement in cross-cultural psychology: a review and comparison of strategies, J. Cross Cult. Psychol. 16 (1985) 131–152.
[10] R.K. Hambleton, P.F. Merenda, C.D. Spielberger, Adapting Educational and Psychological Tests for Cross-Cultural Assessment, Taylor & Francis, City, 2004.
[11] M.J. Gladstone, G.A. Lancaster, A.P. Jones, K. Maleta, E. Mtitimila, P. Ashorn, R.L. Smyth, Can Western developmental screening tools be modified for use in a rural Malawian setting? Archives of Disease in Childhood 93 (2008) 23–29.
[12] A. razavieh, S. shahim, Persian Versian of Wechsler Preschool and Primary Scale of intelligence(WPPSI), shiraz university, City, 2012.
[13] S. Hasanzade, A. Minaei, Persian Version of the Test of Language development.primary(TOLD-P:3). City: Ministry of Education, Research Institute Of Education, 2002.
[14] R. Mohamadi, M.R. Alavije, A. Minayi, Y. Modaresi, M.K. Dastjerdi, M. Ghaderi, Generation and content validation of a Persian syntax comprehension test, 19 (2015) 222.

[15] R.F. DeVellis, Scale Development: Theory and Applications, Sage, City, 2003.
[16] C. Ayre, A.J. Scally, Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation, Meas. Eval. Couns. Dev. 47 (1) (2014 Jan 1) 79–86.
[17] R. Kaplan, D. Saccuzzo, Psychological Testing: Principles, Applications, and Issues, Cengage Learning, City, 2004.
[18] H.K. van der Lely, Language and cognitive development in a grammatical SLI boy: modularity and innateness, Journal of Neurolinguistics 10 (1997) 75–107.
[19] C. Marshall, T. Marinis, Hk van der Lely, Passive verb morphology: the effect of phonotactics on passive comprehension in typically developing and Grammatical-SLI children, Lingua 117 (2007) 1434–1447.
[20] D.V.M. Bishop, P. Bright, C. James, S.J. Bishop, H.K. Van der lely, S.L.I. Grammatical, A distinct subtype of developmental language impairment? Appl. Psycholinguist. 21 (02) (2000) 159–181.
[21] L.B. Leonard, P. Deevy, M.E. Fey, S.L. Bredin-Oja, Sentence comprehension in specific language impairment: a task designed to distinguish between cognitive capacity and syntactic complexity, J. Speech Lang. Hear. Res. 56 (2) (2013) 577–589.
[22] H.K. van der Lely, Canonical linking rules: forward versus reverse linking in normally developing and specifically language-impaired children, Cognition 51 (1) (1994) 29–72.
[23] H.K. Van der Lely, K.J. Heather, L. Stollwerck, Binding theory and grammatical specific language impairment in children, Cognition 62 (3) (1997) 245–290.
[24] M. Silveira, Specific Language Impairment (SLI) Revisited: Evidence from a Psycholinguistic Investigation of Grammatical Gender Abilities in Brazilian Portuguese-speaking Children, PhD thesis University College London, London, 2011.
[25] J.S. Hyde, M.C. Linn, Gender differences in verbal ability: a meta-analysis, Psychol. Bull. 104 (1) (1988 Jul) 53.